IV. Implications for Biology and Society

Biology and the Genome Project

Bob Moyzis: Everyone emphasizes the biotechnology and biomedical spin-offs from this project, but we're also creating an enormous resource for fundamental science, that is, for addressing some of the big open questions in biology.

Gene expression is one of those areas. For the information in the protein-coding sequence of a gene to be expressed as a protein, the DNA sequence is first transcribed into an RNA molecule and then the RNA sequence is translated into a protein. But genes are not actively making proteins all the time. What turns the gene on and off? Most of our models are based on experiments with organisms like *E. coli* where genes are either on or off. The rules of gene expression are much more complicated in humans. Some genes are expressed only in certain tissues and in varying amounts. Some are turned on at one point in development, then turned off again, and then re-expressed in another tissue.

The regulatory networks must be incredibly complicated, but at present we don't have the faintest idea how the expression of even a single human gene is regulated. We've identified some sequences near the gene that we know are important—promoter sequences, for example, that must bind to a special regulatory protein before the gene can be turned on—but some regulatory signals may be very far from the genes themselves. When an individual human gene is put back into a cell in culture, it does not exhibit the exact kind of regulatory patterns as those observed in normal cells. Clearly we have a lot to learn.

In the future you'll be able to dial various sequences into a computer, and from those "area codes" and "molecular addresses" the computer will give you the 10,000 genes that are expressed in the kidney, for example, by virtue of those molecular addresses.

Lee Hood: Once we have the sequences of the human genome, I expect we will find a lot of interesting patterns in the DNA that have to do with gene regulation. Indeed, those will constitute molecular addresses, which tell us in which cells and at what developmental stages the corresponding genes are expressed. We're going to have to figure out ways of deciphering those molecular addresses.

As another example, proteins have to bind to DNA to get it to coil and supercoil into a compact chromosome. I would guess that the DNA sequences that bind to these proteins will be made evident by a detailed analysis of the sequences in complete chromosomes. In the long term, I expect we'll be able to identify the regulatory sequences, that is, the binding sites of proteins, which turn genes on and off, and from those sequences we will deduce where that gene is expressed, when during development it's expressed, and the amplitude of its expression.

The regulatory site will be like a telephone area code, and in the future you'll be able to dial various sequences into a computer, and from those *area codes* and *molecular addresses*, the computer will give you the 10,000 genes that are expressed in the kidney, for example, by virtue of those molecular addresses.

Regulation is probably best understood in *E. coli*, which, of course, is a prokaryote, a cell with no nucleus. In the human, we know a lot about a few elements that regulate some genes, but not many. Extensive studies of gene regulation have been done on bacteria, the sea urchin, yeast, and *Drosophila*. But as far as understanding how a whole genome is put together so that the genes are expressed at the right time, in the right order, and in the right amounts, we have very little information.

David Galas: The global functioning on a whole chromosome won't be understood until we learn how the chromosome is organized not only in terms of the one-dimensional layout of the genes but also in terms of its three-dimensional structure. The basic structural unit of the chromosome is chromatin, which is a double loop of DNA wound around a protein center.

And many of us believe the structure of chromatin plays an important role in gene regulation.

The only things selected by natural selection are the protein products and whether they are turned on and off at the right time. That choreography must involve the detailed structure of the chromosome and how it winds and unwinds during the cell cycle. So if the Genome Project is really an effort to understand all the information encoded in the human genome, then as time goes on, the interests of the Genome Project will become closer and closer to those of structural biology, to the interplay between three-dimensional structures and biological functions.

The sequence information will be relevant not only to chromosome structure but also to protein structure. As we sequence cDNAs and thus determine more protein sequences, we hope to begin to understand how the primary sequences of proteins lead to the three-dimensional structures of the protein macromolecules themselves and therefore to their functions. I say *hope* because we don't know how a sequence of amino acids folds up into a stable protein structure. The protein-folding problem is indeed one of the great conundrums of modern biochemistry and biophysics. Nobody knows whether the problem has a real solution.

If, at the very least, we knew all the protein sequences and all the protein structures, we could figure out how they relate to each other. We don't know whether all the proteins are made up of a relatively small set of little structures, such as alpha helices and beta sheets, or whether each of the 100,000 different proteins is a distinct structure. Maybe there are only 500 elements or modules that are put together in different ways and the various combinations give rise to all the existing proteins.

Bob Moyzis: Theoretically, you can show there hasn't been enough time since the universe began to create at random all the kinds of potential proteins that could be out there. Maybe all that evolution has done has been to mix and match a few hundred basic structural elements to make all the proteins we have. As we determine the sequence of more and more genes, that question will get answered.

David Galas: That idea of mixing and matching a few basic subunits proved relevant to the problem of how our immune system is able to generate a seemingly infinite variety of antibodies in response to foreign invaders. The antibody-diversity problem was solved by posing the existence of combinatorial rearrangements of a relatively small number of subunits with small variations added on here and there.

That explanation seems to have been borne out and provides a particularly elegant, almost mathematical solution to what seemed an almost unsolvable problem about fifteen years ago. The problem of gene regulation may have a similar solution. We may discover a small class of enhancers and promoter regions that form a hierarchy, a computer-program-like structure that governs regulation.

Bob Moyzis: At the risk of sounding like a broken record, I will point out again that this problem of regulation is relevant even to those primarily interested in human disease. Some types of thalassemia, which is the absence of a particular globin protein, are clearly caused by defects in the regulatory region that tells this gene whether or not to be expressed.

David Galas

What we're looking at in the human genome is a historical product of millions of years of evolution . . . . So evolutionary understanding is an inevitable consequence of the Genome Project. You could even characterize the Project as studying evolution.

A working hypothesis is that there are a limited number of *master* genes controlling regulation. The homeobox genes now being intensively studied in *Drosophila* development fall into that category. Their protein products are capable of binding to many different regions of DNA and regulating the expression of genes responsible for the structural development of an organism. And homeobox genes seem to have been conserved through evolution.

Of course the speculation that there are master genes just pushes the basic problem back one more level to how the master switches might be regulated. Again we have no answers, but this is an incredible time to be in biology because with the current explosion in biological knowledge, one has the feeling that we may solve many of these problems within our own lifetimes.

David Galas: What we're looking at in the human genome is a historical product of millions of years of evolution. The more detail we know about the human genome and the genomes of other species, the more we're going to understand about what processes were involved in getting us where we are. So evolutionary understanding is an inevitable consequence of the Genome Project. You could even characterize the Project as studying evolution.

We already know that molecular processes, the control of individual genes, and the structures of mammalian and bacterial viruses are Rube Goldberg-like arrangements. The reason for this seemingly ad hoc complexity is that these organisms and processes developed over time by natural selection and random variation.

At the heart of understanding evolution is understanding developmental control—and that means gene control, turning on a battery of genes at one time versus another.

Bob Moyzis: It's astounding to realize that the tools we're developing to unravel the information content of the human genome will allow us to investigate the DNA from ancient tissues. We'll be able to choose specific STS markers and apply the PCR to very small DNA samples preserved in the bones of our ancestors. It will be like going back in a time machine and finally pinning down some of the speculation about human origins. The more markers we get, the more we are going to be able to answer questions like, where did Cro-Magnon man really come from? People have now isolated DNA from ten-thousand-year-old human samples. This work is almost like science fiction.

We'll be able to choose specific STS markers and apply the PCR to very small DNA samples preserved in the bones of our ancestors. It will be like going back in a time machine and finally pinning down some of the speculation about human origins.

David Galas: The plant record is much older. There's a lake in Minnesota where some leaves have been preserved in an anoxic sediment that dates back to 20 million years ago—they've gotten DNA sequences from the rubisco genes of magnolias and sycamores. Those sequences are much the same as they are today. But there must be some other interesting DNA fossils in that sediment. This is a whole new area of exploration. In terms of evolutionary development, we're farthest along in understanding insect morphologies. Certain classes of Arthropoda have identical segments, like centipedes. Later in evolution came batteries of homeobox genes that caused differences to occur among those segments. Homeotic means changing, and homeotic mutations are those that change parts of the organism by changing individual segments—say, by making the second thoracic segment into an abdominal segment. In some cases you can take identical segments, lay on another level of genetic control, and produce a difference between the segments.

Clearly that's what has happened in the evolutionary branching among those various sorts of Arthropoda. We know that in the early embryonic development in *Drosophila* there are three or four genes, so-called segmentation genes, that lay down the initial segmented pattern of the organism. Then other genes turn on to produce changes among the segments. We're beginning to work out that circuitry now.

Bob Moyzis: Nobody on the Human Genome Project ever talks about development because the reality is that human development is particularly difficult to study.

David Galas: That's an important point. There are very important outstanding questions in development that can be answered in the nematode and in *Drosophila* and so forth. But nematode development, for example, is hard-wired—you know where every cell

goes. Every organism that develops properly through its time cycle with the same genome is identical. It's got the same cells in exactly the same position. That's not true for organisms even a little farther up the scale like *Drosophila*.

When you get to organisms with relatively complex brains, a large part of development is ultimately determined by the particular genome. It's completely stochastic relative to the kind of programming that appears in nematodes. For a decade or more people have been collecting genes in *Drosophila*, and it's only been very recently that they're starting to understand a little about how the genes are controlled relative to one another. That understanding is built up through controlled experiment.

Bob Moyzis: Obviously we're not going to be doing controlled experiments on human development, but we can work with mice, which are similar to humans in many ways. The mapping and sequencing of the mouse genome is part of the Genome Project. And once we have those tools, we can target genetic changes in mice that will give us clues about developmental questions.

The DOE has a history of being interested in agents that cause abnormalities in development, agents that alter the expression of a particular gene and thereby produce an abnormal embryo. The Oak Ridge people, for example, have a really nice set up for making transgenic mice, and they've been able to identify a number of interesting developmental genes in the mouse that have human homologues. And I believe that work will increase as a spin-off from the Human Genome Project. The Human Genome Project is focused on humans, but we need to study a lot of other organisms to understand human development and pathology.

Ethical, Legal, and Social Implications

Bob Moyzis: The Genome Project will have many practical consequences for society and maybe we should close this discussion by addressing some of them.

David Galas: Both the NIH and the DOE are devoting 3 percent of their total Genome Project budgets to the task of addressing the Ethical, Legal, and Social Implications [ELSI] of the public use of genetic information. [The NIH recently increased their allocation to 5 percent.] The ELSI working group was established by the NIH to identify the most pressing issues and to find ways to help make the new information a real benefit to society.

The basic problems . . . are not new—they will simply be exacerbated.

Nancy Wexler: ELSI is a very exciting aspect of the Genome Project. Traditionally, social issues and scientific work have been viewed as separate realms—scientists go into their labs, do their work, and when they finish, it's it up to society to take it as it comes. The Genome Project is different. Scientists like David Galas and Jim Watson recognized *up front* the need to pay attention to the social and ethical implications of their work, and by funding ELSI as an integral part of the Genome Project, they are taking responsibility for the initial examination of the effects it will have on our society.

David Galas: There are two important things to remember when we think about ethical and social issues in terms of the Genome Project. First, there are no *new* problems. Issues concerning privacy, confidentiality, and discrimination will become much more pressing once the Genome Project generates the tools to diagnose genetic diseases presymptomatically. The *basic* problems, however, are not new—they will simply be exacerbated.

The second thing to keep in mind is that many ethicists, lawyers, and social scientists who speak out about the implications of the Genome Project are often somewhat ignorant of the fundamental science of genetics. We need everyone to learn and understand the difference between being a carrier of an abnormal gene and having a genetic disease, between the genetic markers for a disease and the disease gene itself, and between genetic probabilities and genetic certainties. Often, without the benefit of a solid background in genetics, people tend to adopt the attitude of naive genetic determinism, that there are good genes and bad genes or that genes alone control behavior. Those misunderstandings have been around a long time, and we have to start dealing with them.

Nancy Wexler: Education of both the general public and professional healthcare providers is among ELSI's high-priority goals. We are actively encouraging the leaders of voluntary health organizations and genetic-disease support groups to participate in public discussions geared toward creating a greater understanding of the nature of genetic disorders and the issues surrounding the Genome Project.

We are also encouraging the insurance industry to anticipate the challenges they will face as vast quantities of new genetic information become available to the public. In a way, the Project is quite a nuisance to the insurance providers.

UNRAVELING THE CHROMOSOME

E. Morton Bradbury

Central to biology is an understanding of the organization, structure, and functions of the chromosomes of higher organisms. Chromosomes contain the DNA molecules of the genome and are themselves contained within the cell nuclei of all eukaryotes, from single-celled yeast all the way up the evolutionary ladder to human beings. As pointed out by David Galas (pages 164–165 of "Mapping the Genome"), to understand the functions of the multitude of protein-coding and noncoding DNA sequences that will be determined by the Human Genome Project, we will need detailed knowledge of the three-dimensional structure of chromosomes and the structural changes that chromosomes undergo during the various phases of the cell cycle. Major advances in biology will be at the interfaces between the Human Genome Project, structural biology, and molecular biology of the cell.

The size of the human genome suggests the magnitude of the problem. The diploid human genome contains 6×10^9 base pairs or 204 centimeters of DNA molecules packaged into 46 chromosomes. It is generally believed that each chromosome contains a single DNA molecule several centimeters in length.

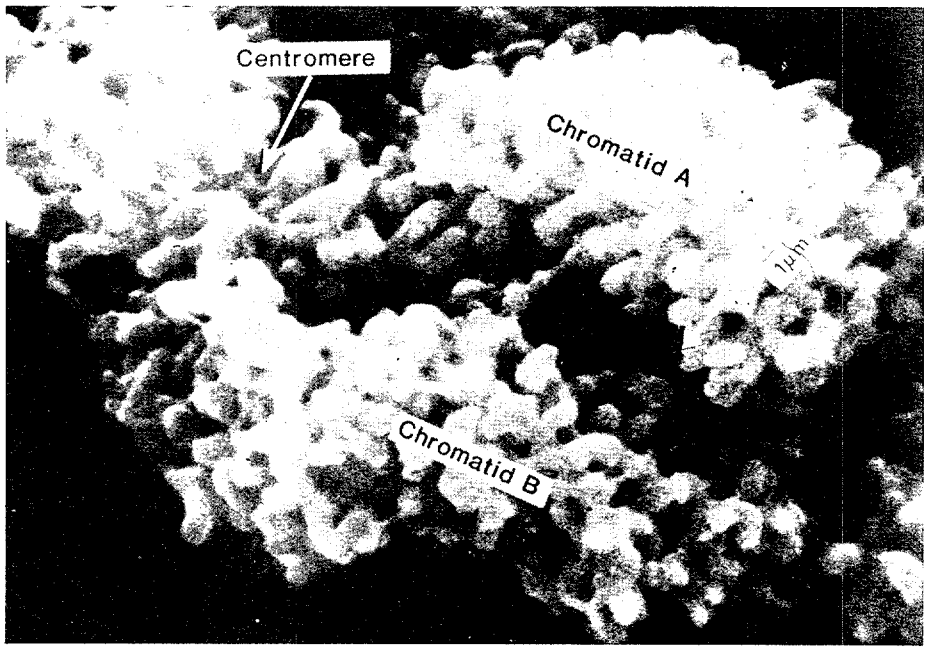


Figure 1. Human Metaphase Chromosome
A scanning transmission electron micrograph of a metaphase chromosome showing two sister chromatids attached at the centromeres. Each compact projection is thought to be a long loop of DNA (see Figure 2) packaged along with various proteins into a thick chromatin fiber. (Reprinted courtesy of U.K. Laemmli, Université de Genève.)

Studies of the yeast *S. cerevisiae*, a lower eukaryote that can be easily manipulated, have revealed three chromosomal elements that are essential to the faithful replication of each chromosome and to the subsequent separation of the two duplicate chromosomes into daughter cells during cell division. These are: (1) the very ends of chromosomes, called the telomeres; (2) a central region of constriction called the centromere that, after replication of a chromosome, is the last point of attachment between the resulting pair of sister chromatids; and (3) a DNA sequence required to initiate DNA replication, called an origin of replication.

Figure 1 is a scanning transmission electron micrograph of a human metaphase chromosome, the highly condensed structure adopted by the chromosome during metaphase, one of the last phases of cell division. The chromosome has already replicated into two sister chromatids. The centromere connecting the sister chromatids (seen in the micrograph as a region of constriction) provides the point of attachment

for the spindle apparatus that contracts and separates the replicated chromosomes into the daughter cells. The telomeres at the ends of each chromatid contain tandem repeated DNA sequences that cap, protect, and maintain the linear DNA ends of the chromosomes during replication.

Each of the 46 human chromosomes can be identified during metaphase by its length, the location of its centromere, and the particular banding pattern produced by staining the DNA of that chromosome. (Banding patterns can be seen in "Chromosomes: The Sites of Hereditary Information" in "Understanding Inheritance.") The origins of the distinctive banding patterns are not well understood but probably reflect a reproducible pattern of DNA folding induced by DNA-protein interactions specific to each chromosome. The DNA molecule is very tightly wound during metaphase. For example, human chromosome 16 is 2.5 micrometers long, whereas the DNA molecule in each sister chromatid is 3.7 centimeters long. In other words, the packing ratio of the linear DNA molecule in the metaphase chromosome is 15,000 to 1.

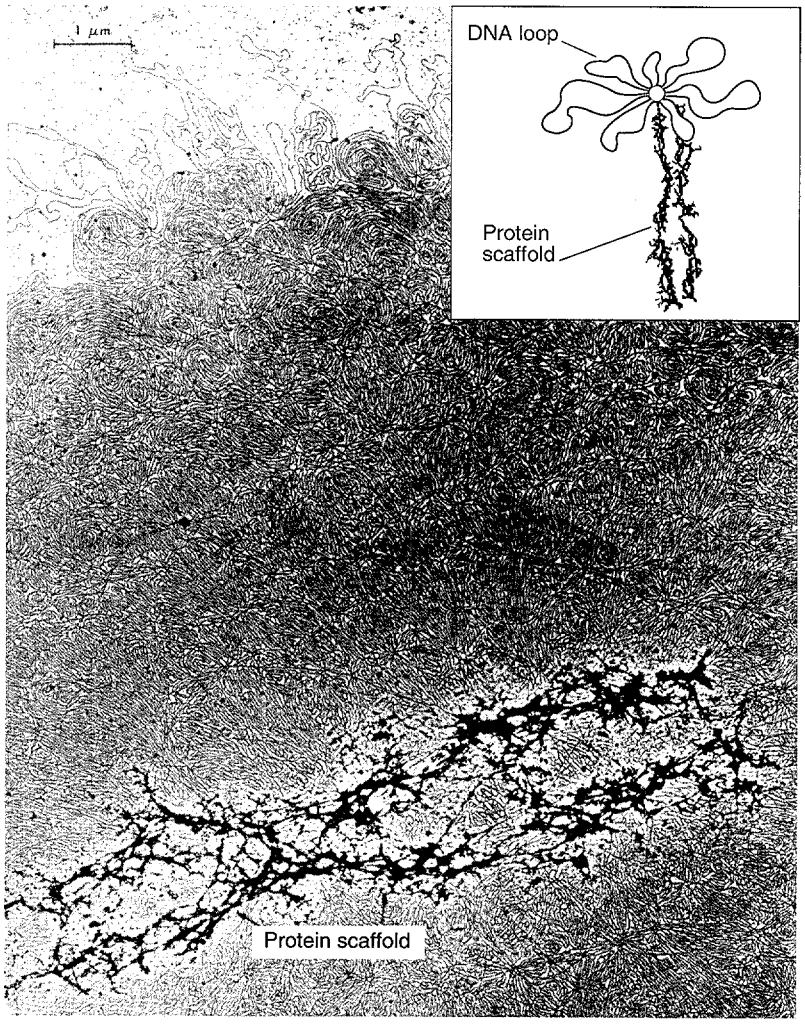


Figure 2. Chromosome Loops and Protein Scaffold

Above is a metaphase chromosome depleted of almost all chromosomal proteins. The remaining 2 to 3 percent of the proteins form a scaffold that retains the shape of the intact chromosome. Around the scaffold is a halo of loops of naked DNA. Each loop appears to begin and end at the same point along the protein scaffold (see insert). The number and sizes of these loops suggest that each may contain a single gene or a group of linked genes. (Reprinted courtesy of U.K. Laemmli, Université de Genève.)

Chromosomal DNA Loops

When chromosomal material is isolated from the nucleus, the long DNA molecules are found to be associated with chromosomal proteins, whose weight is up to twice that of the DNA. The five histones, the many copies of which are equal in weight to that of DNA, are found in all eukaryotes and as explained below are involved in packaging the DNA in the chromosomes. The non-histone proteins are a heterogeneous group and many are associated with the various chromosome functions, such as replication, gene expression, and chromosome organization. Among the latter are a small group that bind most tightly to the DNA and form a scaffold for the chromosome. This protein scaffold has been made visible by gently treating metaphase chromosomes with detergents to remove the histones and most other nonhistone proteins. The remarkable structure that remains is shown in Figure 2. The residual protein scaffold, or "ghost," of the metaphase chromosome is surrounded by a halo of DNA. At higher

resolution DNA loops can be observed to emerge from and return to the same point on the protein scaffold (see inset in Figure 2).

Two major scaffold proteins have been isolated, Sc1 and Sc2. Sc1 has been identified as topoisomerase II, an enzyme that relaxes supercoiled DNA by cutting through both strands of the DNA, thereby enabling the cut DNA ends to rotate, and then resealing the cut. The cuts made by topoisomerase II are essential for the separation of sister chromatids to the daughter cells.

The DNA loops in Figure 2 range in size from 5,000 to 120,000 base pairs and have an average size of about 50,000 base pairs. Thus the haploid human genome of 3×10^9 base pairs of DNA corresponds to 60,000 loops, which is close to the estimated numbers of genes, 50,000 to 100,000, in the human genome. Perhaps each DNA loop contains one or a small number of linked genes and therefore serves as both a genetic and a structural unit of eukaryotic chromosomes. This tantalizing conjecture was first made in 1978, and although it remains unproven, evidence in its favor has been accumulating.

Chromatin Contains a Repeating Subunit Structure

Having looked at some of the largest structural features of the chromosome, we now turn to what we know about the small, repeating substructures within a chromosome. DNA with its associated chromosomal proteins, histones, and nonhistone proteins, is called chromatin. In 1973 chromatin in isolated nuclei was first digested with micrococcal nuclease, an enzyme that cuts double-stranded DNA. The digestion yielded a ladder of DNA lengths in multiples of about 190 to 200 base pairs. Evidently DNA sequences spaced by 190 to 200 base pairs were more accessible to attack by micrococcal nuclease than the intervening DNA. This seminal observation showed that chromatin contained a simple, repeating subunit, known as the nucleosome.

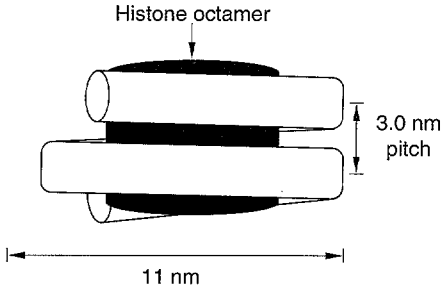


Figure 3. Nucleosome Core Particle
Structure of the nucleosome core particle determined from neutron scattering. The core particle is a flat disc, 100 angstroms in diameter and 55 to 60 angstroms thick.

For most somatic tissues, the nucleosome contains three elements, a stretch of DNA containing 195±5 base pairs, one copy of the histone octamer $[(H3_2H4_2)(H2A,H2B)_2]$ and one copy of the histone H1. More prolonged micrococcal nuclease digestion reduces the length of the DNA in the nucleosome, thereby creating a slightly smaller unit, called the chromatosome, which contains 168±2 base pairs of DNA, the histone octamer, and H1. Such digestion often reduces the nucleosome to an even smaller unit contained within the chromatosome and called the nucleosome core particle. It contains 146±1 base pairs of DNA and the histone octamer (see Figure 3).

The nucleosome core particle has been obtained in large quantities and subjected to extensive structural studies. In 1974 neutron-scattering studies of the core particle in aqueous solution showed that it was a flat disc of diameter 100 angstroms and thickness 55 to 60 angstroms, with 1.7 turns of DNA coiled on the outside of a core of the histone octamer at a pitch of about 30 angstroms

(Figure 3). Subsequent x-ray-diffraction studies of crystallized core particles achieved a resolution of 6 to 7 angstroms. The crystal structure (Figure 4) not only confirmed the lower resolution solution structure achieved by neutron scattering but also showed that histones are in contact with the minor groove of DNA and leave the major groove available for interactions with the proteins that regulate gene expression and other DNA functions. The 7-angstrom-resolution crystal structure also revealed that DNA does not bend uniformly but rather bends gently and then more sharply around the histone octamer. Such a path implies that flexibility, or bendability, of DNA may be sequence-dependent and that the underlying DNA sequence along the molecule may determine the positions of some nucleosomes. The most recent work on nucleosome positioning shows that the bulk of nucleosome core particles are able to move along the DNA molecule between a cluster of positions separated by about 10 base pairs. This mobility is probably required during DNA replication and transcription to allow DNA polymerases and other enzymes access to specific DNA sequences.

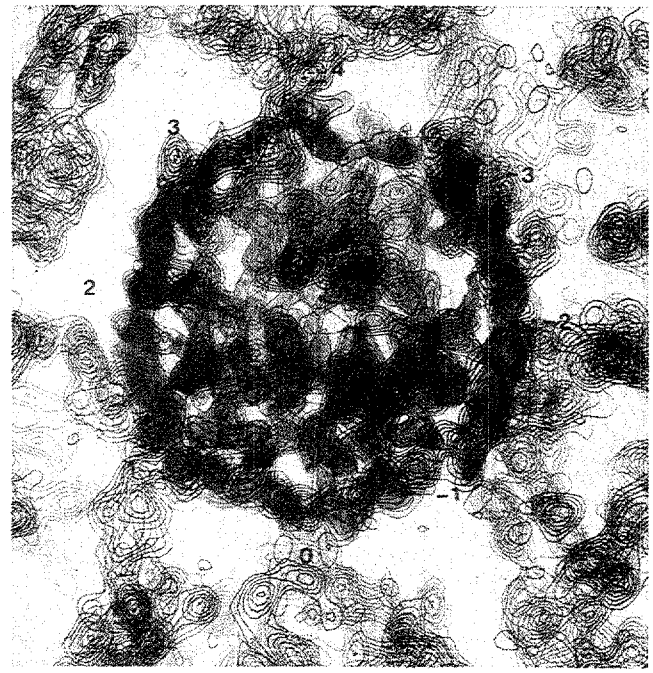


Figure 4. Crystal Structure of Core Particle

The structure of the nucleosome core particle as determined by x-ray diffraction is shown above. At a resolution of 6 to 7 angstroms, this top view of the core particle shows that the DNA (brown) does not follow a smooth path around the histone octamer (blue and turquoise) but rather bends sharply and then more gently. (Reprinted courtesy of Uberbacher and Bunick, Oak Ridge National Laboratory.)

Despite considerable effort to achieve higher resolution, the best data for the core particle structure is at a resolution of about 6 angstroms. However, the crystal structure of the isolated histone octamer has been solved to the higher resolution of 3.3 angstroms. This structure shows shapes of the individual histones and the nature of interhistone interaction of most but not all of the histone polypeptide chains. In particular, the basic N-terminal domains, comprising some 20 to 25 percent of the histone octamer, are not "seen" in the crystal structure, probably because they bind to DNA, and in the absence of DNA, they are disordered. These N-terminal domains contain all of the sites of the cell-cycle-dependent acetylation of lysines and phosphorylation of serines or threonines. Acetylation of lysine converts it from a positively charged residue, which can therefore bind to DNA, to a neutral acetylysine. It has been shown first that lysine acetylation is strictly correlated with transcription and DNA replication, and more recently, that histone acetylation drives the uncoiling of part of the DNA from the nucleosome to allow the initiation and progression of DNA replication and transcription.

Chromatosomes and Nucleosomes

A model of the structure of the chromatosome (Figure 5) has been inferred from the structures of the nucleosome core particle and the histone H1. The core particle has 1.7 turns of DNA at a pitch of 3.0 nanometers (30 angstroms) coiled around the histone octamer. Consequently, the chromatosome's 168 base pairs of DNA are long

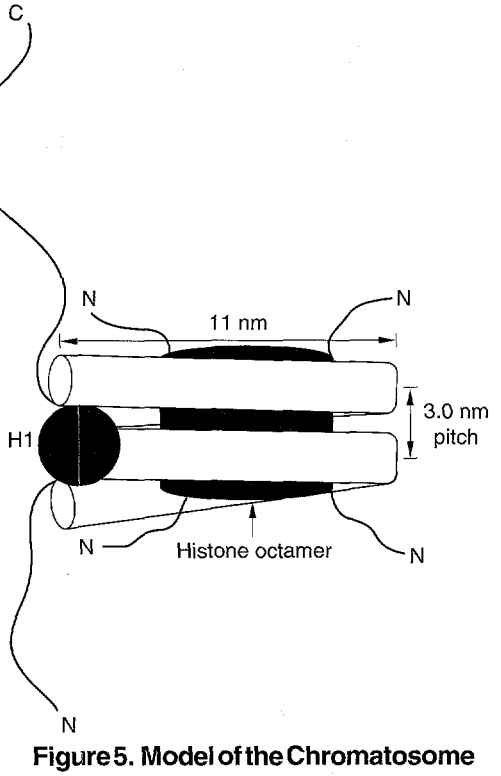


Figure 5. Model of the Chromatosome
The model includes the nucleosome core particle, an extra stretch of DNA, and the histone H1. The DNA makes two complete turns around the histone octamer, and H1 is bound to the outside of the coil at the place where the coil begins and ends. In this position H1 might server to modulate long-range interactions that modify chromosome structure during the cell cycle.

enough to complete two turns of DNA around the histone octamer. The chromatosome also includes the fifth histone H1. In the model structure shown in Figure 5, the histone H1 is bound to the outside of the coiled DNA where it might serve to modulate long-range interactions associated with reversible changes in chromosome structure during the cell cycle. During cell division chromosomes become more and more condensed until they reach metaphase. Then, when cell division is completed and the daughter cells enter interphase, the chromosomes assume a less-condensed configuration (see "Mitosis" in "Understanding Inheritance"). The long, flexible "arms" of H1 undergo a pattern of phosphorylations through this cycle, which may well modulate the long-range interactions required to coordinate these structural changes in the chromosomes. In support of this hypothesis is the fact that an increase in H1 phosphorylation has been correlated with the process of chromosome condensation to metaphase chromosomes. To describe the nucleosome beyond the model for the chromatosome requires a knowledge of the paths of the DNA that link one nucleosome to another. Our present lack of knowledge about those paths impedes our ability to pin down the higher-order chromatin structures that make up the chromosome.

Higher-Order Chromatin Structures

Although higher-order structures of chromatin cannot be resolved in the chromosome itself, they can be studied in solution. Chromatin, when placed in low ionic strength, 10-millimolar NaCl, forms a 10-nanometer-diameter fibril of nucleosomes, which is sometimes referred to as "beads on a string." This form is also observed when chromatin spills out of lysed nuclei. Neutron-scattering studies of the 10-nanometer chromatin fibril give a mass per unit length equivalent to one nucleosome per 10±2 nanometers of fibril, or a DNA packing ratio of between 6 and 7 to 1. When ionic strength is increased to 150-millimolar NaCl, corresponding to normal physiological conditions, the 10-nanometer fibril undergoes a transition to the "30-nanometer" fibril. Neutron-scattering studies indicate that the diameter for this fibril in solution is 34 nanometers and the mass-per-unit length is equivalent to 6 to 7 nucleosomes per 11 nanometers of fibril, or a DNA packing ratio of between 40 and 50 to 1. Figure 6 shows the simplest model of the 34-nanometer fibril that is consistent with available structural data: it is a supercoil or solenoid of 6 to 7 radially arranged disc-shaped nucleosomes with a pitch of 11.0 nanometers and a diameter of 34 nanometers. Basic questions concerning the location of histone H1 and the linker DNA connecting the nucleosomes remain unanswered.

Packaging of Chromosome Loops

With these higher order chromatin structures in mind, we can imagine how the large transverse DNA loops present in the histone-depleted metaphase chromosome (see Figure 2) might be packaged in the normal chromosome. Since the average size of the

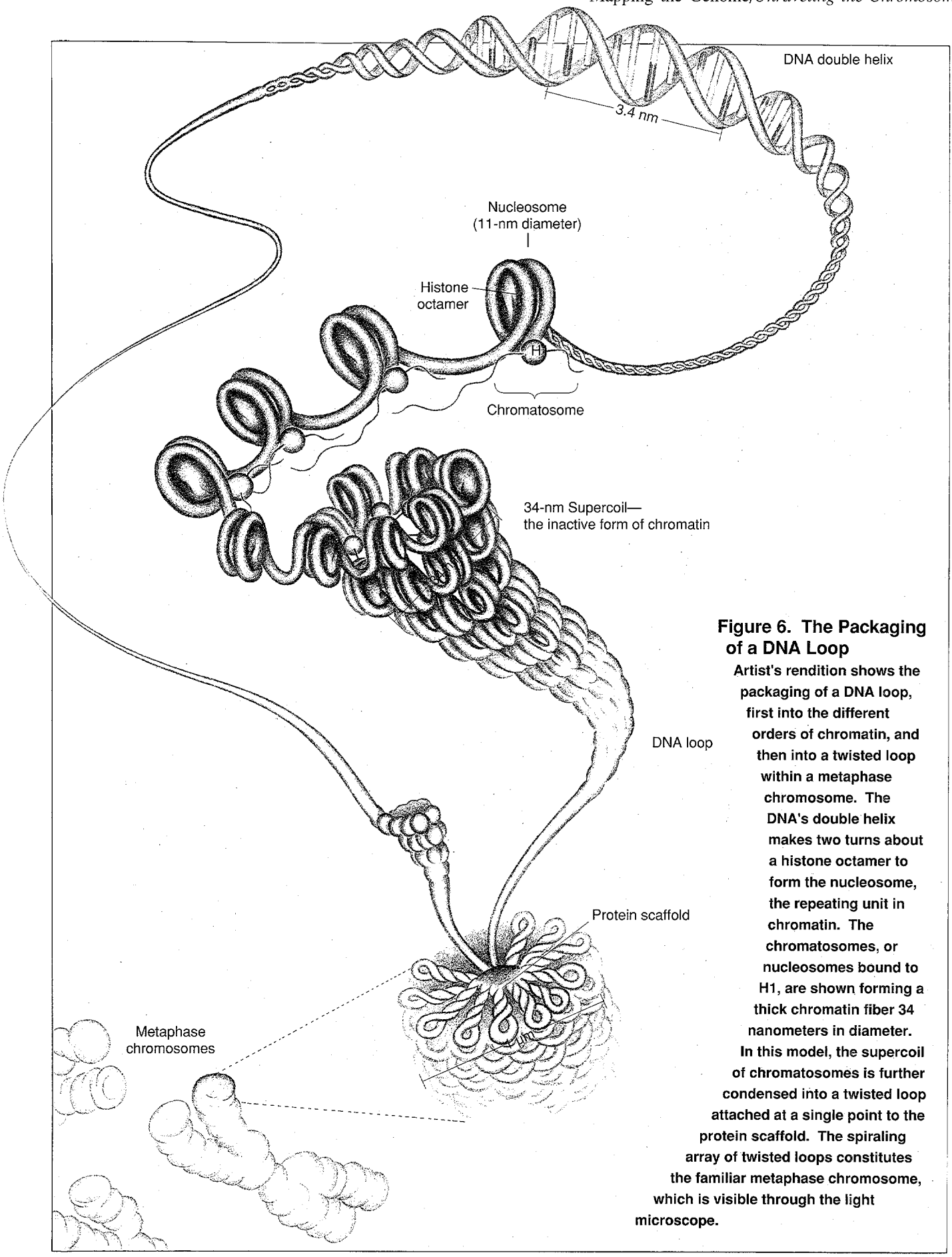


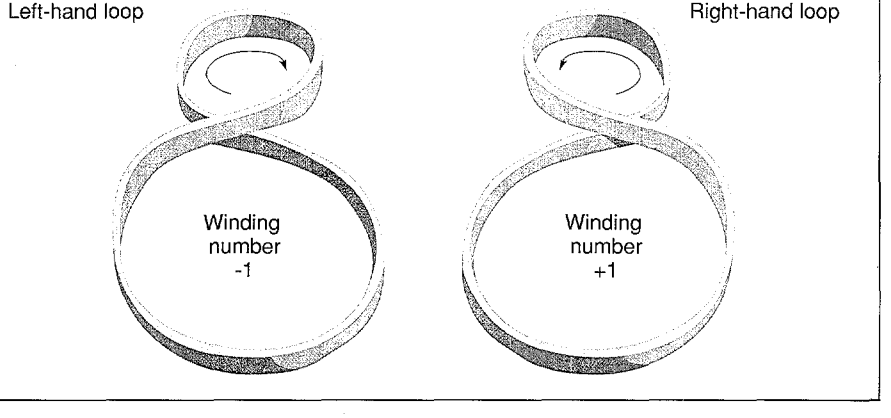
Figure 6. The Packaging of a DNA Loop

Artist's rendition shows the packaging of a DNA loop, first into the different orders of chromatin, and then into a twisted loop within a metaphase chromosome. The DNA's double helix makes two turns about a histone octamer to form the nucleosome, the repeating unit in chromatin. The chromatosomes, or nucleosomes bound to H1, are shown forming a thick chromatin fiber 34 nanometers in diameter. In this model, the supercoil of chromatosomes is further condensed into a twisted loop attached at a single point to the protein scaffold. The spiraling array of twisted loops constitutes the familiar metaphase chromosome, which is visible through the light microscope.

DNA loops is 50,000 base pairs, or 17 micrometers in length, each loop of DNA can form a string of nucleosomes that are either coiled to form 2.6 micrometers of a 10 nanometer fiber, or supercoiled into 0.4 micrometers of a 34 nanometer fiber. Thus, to create the thickness of a sister chromatid (Figure 1), which is 1 micron in diameter, would require just one more order of chromatin folding above the 34 nanometer supercoil. Figure 6 shows a possible model of this final level of chromatin folding.

How is the packaging of DNA loops controlled in response to chromosome functions? Evidence suggests that the inactive form of chromatin is the 34-nanometer supercoil or solenoid of nucleosomes. For both DNA transcription and genome replication this supercoil of nucleosomes must first be uncoiled to the linear array of nucleosomes and then the DNA must uncoil even further to allow access of the transcriptional machinery or the replication machinery to the DNA sequences. Whenever DNA is constrained by proteins to form a loop, DNA supercoiling becomes an important consideration in understanding DNA structure-function relationships. DNA supercoiling has been subjected to extensive experimental and mathematical analysis.

Figure 7. Configurations of a Closed Loop of Ribbon with Winding Numbers +1 and -1



Consider a model in which each DNA loop is firmly attached to the protein scaffold of a chromosome and is therefore somewhat analogous to a closed loop of ribbon. A closed loop of ribbon has a topologically invariant property known as the winding number, which is the number of twists in the ribbon plus the number of times the ribbon crosses itself, that is, coils about itself. The winding number is an integer or half-integer and remains constant unless the ribbon is cut. Each complete twist and each complete crossing adds $+1$ or -1 to the winding number depending on the direction of the twist or crossing. A right-handed twist (the same direction as the thread of a standard screw and the standard helical structure of a double-stranded DNA molecule) is positive, and a left-handed twist is negative. Similarly, a crossing that produces an extra right-handed loop in a loop of ribbon is positive, and a crossing that produces an extra left-handed loop in a loop of ribbon is negative (see Figure 7).

Now consider a loop of double-stranded DNA. Unconstrained DNA has 10.4 to 10.6 base pairs in each complete turn of the double helix. Taking the value 10.6 base pairs per helical turn, the twist (*Tw*) of a loop of unconstrained DNA consisting of N base pairs would be $N/10.6$. Because a double-stranded DNA molecule already has a helical structure, a loop of DNA further coiled about itself is said to be supercoiled. The linking number (*Lk*) of a closed loop of DNA is defined in terms of the twist and the number of supercoils, or writhe (*Wr*), through the equation $Lk = Tw + Wr$. Twists can be converted into supercoils, but *Lk* must remain constant in a DNA loop whose ends are fixed, in analogy with the constancy of the winding number of the loop of ribbon. If the loop is closed, the linking number must be an integer.

As an example, suppose three helical turns of a linear stretch of DNA are unwound and the ends are then joined. The linking-number change resulting from the unwinding is -3, and the loop can take on any of the three configurations shown in Figure 8. Moreover, the three configurations can be converted into one another without cutting the DNA. DNA configured as in (b) and (c) is said to be negatively supercoiled.

As shown in Figure 3, the DNA in the nucleosome core particle has 1.7 left-handed supercoils and in early studies it was expected that the linking-number change associated with the dissociation of a core particle would be -1.7. However, the experimentally determined linking-number change was -1.02. Although this difference was unexpected and initially controversial, it is easily explained by the change in twist between the DNA constrained in the core particle and free DNA in solution. The average DNA helical repeat on the core particle as measured from its crystal structure is 10.1 base pairs per turn. If we take the average helical repeat of free DNA as 10.6 base pairs per turn, the difference in twist between the DNA in the core particle and free DNA would be $146/10.1 - 146/10.6$ that is, 0.68. Thus the linking-number change associated with the core particle $\Delta Lk = -1.7 + 0.68 = -1.02$ as observed.

Now we can suggest how a DNA loop packaged as a 34-nanometer supercoil of nucleosomes (see Figure 6) could be unwound during interphase. If negative supercoils previously constrained by the nucleosomes are released, then negative supercoiling must be taken up by the linker DNA joining one nucleosome to another. This negative supercoiling would favor the unwinding of a 34-nanometer supercoil of nucleosomes. As suggested above, the acetylation of histones releases DNA that was negatively supercoiled about the histone octamer, presumably by unwinding DNA from the ends of the nucleosome.

The reverse process of chromosome condensation to the metaphase configuration (see Figure 1) requires that the 34-nanometer supercoil be further coiled into higher orders of coiling(s). Perhaps histone-H1 phosphorylation introduces additional supercoiling into a packaged DNA loop causing the higher order of coilings of metaphase chromosomes.

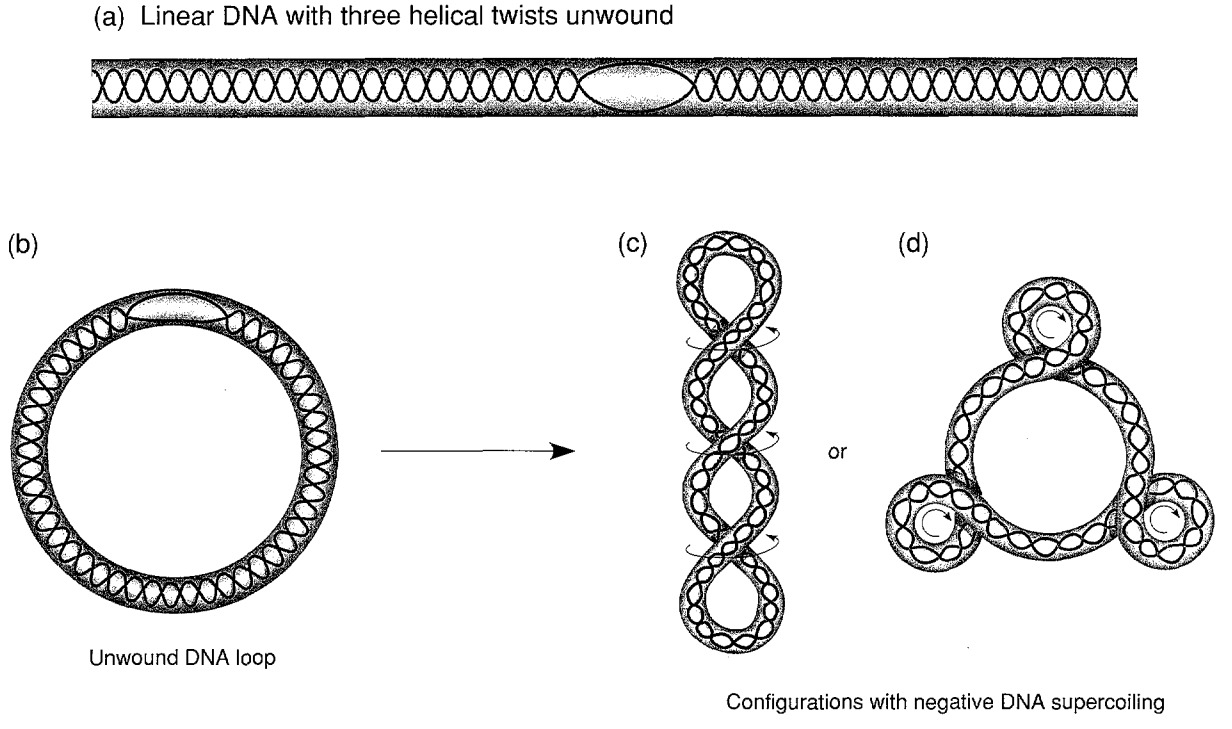


Figure 8. Negative Supercoiling of a Closed DNA Loop
If three helical twists of a linear, double-stranded DNA molecule are unwound as shown in (a) and the ends are then joined, the resulting DNA loop can take on the configurations shown in (b), (c), and (d). All three have the same linking number. In (b) the circular molecule is missing three helical twists that would be present in the normal structure. In (c) the three twists are restored and the loop forms a right-handed superhelix with three crossings in (d) the three twists are restored, but the loop forms three extra left-handed loops. Configurations (c) and (d) are referred to as negative DNA supercoiling.

Figure 8 shows in outline the different orders of packaging of DNA loops into the different orders of chromatin structure and into metaphase chromosomes. It appears that the reversible chemical modifications of acetylation and phosphorylation of histones are involved in the structural transitions undergone by a chromosome during the cell cycle. These structural transitions are dictated by the functional requirements of chromosomes.

Conclusion

Despite recent advances in understanding centromeres and telomeres, we are still a long way from understanding the relationships between structure and function of eukaryotic chromosomes. Relevant to this understanding will be the sequence information from the Human Genome Project. Although much interest is now focused on the mapping and sequencing of genes, the noncoding DNA regions clearly

contain information involved in the organization and functions of chromosomes. The constancy of the banding patterns of individual metaphase chromosomes reflects a highly reproducible pattern of long-range DNA folding, most probably directed by specific DNA-protein interactions and possibly by unusual DNA structures such as bent DNA segments. Superimposed on the very long-range order suggested by banding patterns is the packaging of the DNA loops by the histones together with other structural and regulatory proteins.

The existence of several subtypes of each histone raises the possibility that DNA loops containing different gene families could be packaged with different types of histones according to the requirements of the different cells. DNA control regions of active genes must be packaged in a fashion that makes them accessible to gene-regulating proteins, whereas regions containing permanently repressed genes of a particular cell type may be packaged so that they are inaccessible to such proteins. Such packaging may also determine the availability of DNA regions to chemical damage. Thus a knowledge of the organization of chromosomes is essential to an understanding of the central processes of cell differentiation and the orderly development of complex organisms as well as the processes of DNA damage in chromosomes. ■

Further Reading

E. Morton Bradbury, "Reversible Histone Modifications and the Chromosome Cell Cycle." *BioEssays*, Volume 14, No. 1. January 1992.

Morton Bradbury received a bachelor of science degree in physics and a Ph.D. in biophysics from King's College, University of London, in 1955 and 1958, respectively. After completing his postdoctoral research at Courtauld Research Laboratory, he was appointed head of the Department of Molecular Biology at Portsmouth (England) Polytechnic in 1962, where he remained until his appointment at UC Davis in 1979. He became leader of the Life Sciences Division at Los Alamos in 1988. Bradbury's research has been devoted to understanding whether chromosome organization and chromosome structure are involved in determining how a cell looks and behaves; the structure and function of active chromatin; and the process by which chromosomes condense prior to cell division. In pursuing his investigations, Bradbury has combined the results of measurements derived from the use of a wide range of techniques, including optical spectroscopy, nuclear magnetic resonance, x-ray diffraction, electron diffraction, and neutron diffraction. The recipient of numerous award and honors, Bradbury has also chaired a number of scientific organizations, including the British Biophysical Society, the International Council for Magnetic Resonance in Biology, and the Neutron/Biology Committee of the Institut Laue-Langevin. Bradbury is a member of HERAC and a member of the HERAC subcommittee on structural biology.

Nancy Wexler

Ultimately, the public will have to exert their democratic powers in order to implement the changes that meet their needs. The research and activities funded by ELSI are intended as a catalyst for public discussion of the problems and a foundation for their eventual solutions.

They have already worked out their actuarial tables, and they don't necessarily want to do that all over again. I don't believe that the insurance companies are welcoming future genetic-testing capabilities as a wonderful new tool that will enable them to better discriminate against people, but they need to figure out how the Genome Project will affect their business, what changes they can expect, and how they will handle these changes.

ELSI's Insurance Task Force includes representatives from the insurance industry, corporate benefit plans, and consumer and health groups who are all working together to come up with a plan of action for developing guidelines for insurance policy by 1993. It is significant that all these groups are working *together* in order to arrive at practical, realistic solutions to complex problems.

David Galas: In the past, society has largely ignored problems of genetic discrimination because too few people were affected. But a significant number of people have lost their health insurance or been discriminated against in employment because someone, in ignorance, decided they were at risk for some genetic disease. With the introduction of new genetic tests, more and more people will become vulnerable to such injustices.

Now, because the Genome Project has a very high visibility, we can stimulate society to come to grips with some fundamental problems. We need to address old problems like the confidentiality of medical records, what medical insurance really means in our society, and what it means to spread out risks. These problems are very difficult, and anyone who says they are going to be solved easily is just not thinking.

Nancy Wexler: Certainly, ELSI can't be counted on to *solve* all the problems. What ELSI *can* do is provide an infrastructure or framework for anticipating and describing the basic issues. We have begun to fund research grants, sponsor conferences, provide fellowships, and commission studies in order to stimulate positive changes.

We're trying to pave the way for practical responses to social challenges—old and new—through research, planning, and the development of public awareness. Ultimately, the public will have to exert their democratic powers in order to implement the changes that meet their needs. The research and activities funded by ELSI are intended as a catalyst for public discussion of the problems and a foundation for their eventual solutions.

David Galas: But we need to encourage people to submit proposals dealing with the hard issues that will impact the public directly. I've seen some of the proposals submitted to the DOE by the academic community, and in my view, many of them are unnecessary and rather off the mark. For example, I read one proposal aimed at studying the implications of the Human Genome Project for reductionism. Reductionism is a perfectly fine thing to study, but the Genome Project is not any more reductionist than the rest of biology.

We need studies and proposals that make specific suggestions about the legal agenda, educational programs, and pilot projects that will be really useful. A good example is the NIH-sponsored pilot project to study the problems associated with making a genetic test for cystic fibrosis available to the public. That study will focus on many significant issues such as confidentiality and methods of information delivery and genetic counseling.

David Cox: The pilot project on cystic-fibrosis testing was initially spearheaded by the Human Genome Project. Since the map of the human genome will allow for the isolation of many disease genes and the diagnostic testing of people at risk for those diseases, we believe that the cystic-fibrosis screening program will provide a valuable model for future genetic-screening programs.

Of course the NIH has for many years been involved with ensuring that the results of basic research on diseases are delivered to the public in the form of new medical services. After the cystic-fibrosis gene was cloned, we heard a lot of talk about a testing program, but at that time none of the NIH organizations were willing to put up the money to figure out in advance the best way to deliver screening and counseling to the general public. The community of scientists and healthcare professionals came forward and said, "This is a gap that is not being filled. If you guys say that the Genome Project is going to improve the quality of genetic services in our society, then you had better get on the stick!" And so we did. The Genome Project took the lead in initiating the pilot program, and as a result, the NIH called for applications for research grants to develop the best methods for delivering genetic services.

Nancy Wexler: The cystic-fibrosis pilot studies are being carried out by a group of seven research teams around the country. The research is supported by grants from three different NIH components—the National Center for Nursing Research [NCNR], the National Institute of Child Health and Human Development [NICHD], and of course, the National Center for Human Genome Research [NCHGR]. ELSI will be managing and coordinating the three-year study. All of the groups involved have a strong interest in the ways in which new genetic tests are integrated into clinical practice.

Rather than slow the science, we need to accelerate the creation of a social system that will be more hospitable to new information about our genes, our heritage, and our future. It's a big job, but it's very exciting.

David Cox: Developing methods of screening for cystic fibrosis is complicated by the fact that there are many different mutations of the gene that causes the disease, and at present the tests screen for most but not all of them. First, we have to decide what type of screening is and isn't possible, and then, what services should or should not be made available.

Eventually, the screening technology developed by the Project will be transferred to the commercial sector, but commercialization is a complicated issue. Many companies want to offer cystic-fibrosis screening because such a venture would be quite profitable. However, there are no regulatory systems in place that will ensure that the interests of the public are protected. We have no data that tell us exactly what problems to expect nor have we yet developed methods to address those that might arise.

The American Society of Human Genetics made a statement several years ago recommending against general population screening for cystic fibrosis until more information was available. Their recommendation made some companies very unhappy. Now there is increasing pressure to ignore that recommendation and to proceed with screening in answer to an inferred demand on the part of the community. That's why the pilot projects are so crucial. Without well thought-out regulations based on experience and sound data, every time a gene is cloned there will be a free market where new tests are offered without regard to the impact on the public.

It's clear that we need concrete and practical initiatives, involving both the public and the scientific community, that give suggestions as to how the Genome Project can be most beneficial to society. But suggestions that merely prohibit the main goal of the Project, which is to construct the tools for deciphering the human genome, will not be useful or beneficial. Balancing the long-term vision with more immediately practical concerns is very difficult, but it must be done if the Project is to use this country's resources to best advantage.

Nancy Wexler: The Genome Project is going to have a profound effect on people because it is so closely related to how we function, how we live, how we become ill, and how we heal ourselves. It would be foolish to try to slow down the advancing science—the advances promise better treatments for disease, better quality of life and health for society. Rather than slow the science, we need to accelerate the creation of a social system that will be more hospitable to new information about our genes, our heritage, and our future. It's a big job, but it's very exciting. ■

[For further discussion of ELSI issues, see "ELSI: Ethical, Legal, and Social Implications" and "An Invitation to Genetics in the 21st Century."]

The Participants

David Baltimore received a B.A. in chemistry from Swarthmore College in 1960 and a Ph.D. in biology from Rockefeller University in 1964. In 1975 he shared the Nobel Prize in Physiology or Medicine with Howard Temin and Renato Dulbecco for "discoveries concerning the interaction between tumor viruses and the genetic material of the cell." In 1970, independently of but simultaneously with Temin, he discovered reverse transcriptase, making possible much of the innovative genetic research that followed. Baltimore has served in academic posts including postdoctoral fellow at MIT (1963–1964), postdoctoral fellow at Albert Einstein College of Medicine (1964–1965), research associate at the Salk Institute (1965-1968), associate professor at MIT (1968–1972), and Professor of Biology at MIT (1972–1990). He also served year-long appointments as American Cancer Society Research Professor and as Director of the Whitehead Institute. In July 1990 he became president of Rockefeller University; he resigned from that position in December 1991 but remains on the faculty. Baltimore has received numerous scientific awards and has been a leading spokesperson on many national and international issues related to science including genetic research, biological warfare, AIDS research, and the regulation of science.

David Botstein earned his A.B. from Harvard University in 1963 and his Ph.D. from the University of Michigan in 1967. He then joined the faculty of the Massachusetts Institute of Technology where he eventually became a Professor of Genetics. In 1990 he moved to his present position as Professor and Chairman of the Department of Genetics at Stanford University School of Medicine. Botstein's research has centered on genetics. The bacteriophage P22 was the focus of his earliest work, which included studies of DNA replication, recombination, assembly of the viral head, and DNA maturation. In the 1970s he studied the budding yeast *Saccharomyces cerevisiae* and developed novel genetic techniques to study the functions of the actin and tubulin cytoskeletons. In 1980 Botstein made a significant theoretical contribution to human genetics when he suggested, with collaborators, that restriction-fragment-length polymorphisms (RFLPs) could be used to produce a linkage map of the human genome and to map the genes that cause disease in humans. Botstein has won many scientific awards including the Genetics Society of America Medal (1985), the Allen Award of the American Society of Human Genetics (1989), and the 1992 Rosenstiel Award. He serves on the Advisory Council of the National Center for Human Genome Research and, along with R. W. Davis, is helping to organize the Stanford Yeast Genome Project.

David R. Cox earned his A.B. in biology from Brown University and his M.D. and Ph.D. in medicine and genetics from University of Washington in 1975. Since receiving his doctorate—except for a brief period of medical residency at Yale New Haven Hospital—Cox has been with the University of California at San Francisco. Currently, he is a professor in the Departments of Psychiatry, Biochemistry, and Pediatrics and is Director of the Medical Scientist Training Program at UCSF. Cox has served in several public advisory positions, has been a member of the Biomedical Sciences Study Section and the Mammalian Genetics Study Section at the NIH, and has served on the Scientific Advisory Board of the Genome Data Base at Johns Hopkins University.

David J. Galas earned his B.A. in physics at the University of California, Berkeley, and his M.S. and Ph.D. in physics at the University of California, Davis, and Lawrence Livermore National Laboratory. Before joining the faculty at University of Southern California in 1981, Galas spent four years at the Molecular Biology Department of the University of Geneva, Switzerland. His research interests have included the study of transposition of genetic elements and the study of DNA-protein interactions. He has developed several techniques used in molecular-biology research, including the widely used DNA "footprinting" method, a technique for determining specific DNA sites that interact with proteins and are involved in the regulation of gene transcription. In April 1990 Galas became the Department of Energy's Associate Director for Health and Environmental Research. Major DOE programs for which he is responsible include the Human Genome Project, the Structural Biology Program, the Global Change Research Program, and the Subsurface Science Program. Galas is a member of several federal advisory boards and scientific societies, and he chairs the Biotechnology Research Subcommittee for the Federal Coordinating Council on Science and Technology.

Leroy Hood received an M.D. from the Johns Hopkins Medical School and a Ph.D. in biochemistry from the California Institute of Technology. His research interests have been focused primarily on the study of molecular immunology and biotechnology. The Hood laboratory has played a major role in the development of automated microchemical instrumentation that permits the highly sensitive sequence analysis of proteins and DNA as well as the synthesis of peptides and gene fragments. More recently, Hood has applied his laboratory's expertise in large-scale DNA mapping and sequencing to the analysis of the human and mouse T-cell receptor loci—an important effort for the Human Genome Project. Hood is a member of the National Academy of Sciences and the American Association of Arts and Sciences. He has received numerous awards including the Louis Pasteur Award for Medical Innovation, the ARCS Foundation Man of Science Award for deciphering the message of DNA, and the Albert Lasker Basic Medical Research Award for studies of immune diversity. In 1990 he received the American College of Physicians Award of Distinguished Service for work in the development of instruments used to study modern biology and medicine. Currently, Hood is Bowles Professor of Biology and Director of the NSF Science and Technology Center for Molecular Biotechnology at the California Institute of Technology.

Robert K. Moyzis is Director of the Center for Human Genome Studies at

Los Alamos National Laboratory and is known for his work on human genome organization. His discovery of the human telomere is a landmark in the history of our understanding of chromosome structure and function. In the language of genetics, this sequence $(TTAGG)_n$, means "the end." Its isolation not only provided the necessary end-points of human chromosomes but also supplied the first evidence that unusual DNA structures can have biological importance as a second informational "code." Moyzis leads the physical-mapping effort at Los Alamos and continues to balance his research and administrative responsibilities in the genome center. He serves on numerous committees that oversee the DOE and NIH Human Genome Project, including the DOE Human Genome Coordinating Committee and the joint NIH-DOE Human Genome Advisory Committee. Moyzis received his B.A. in biology and chemistry from Northeastern Illinois University in 1971 and his Ph.D. in molecular biology from the Johns Hopkins University in 1978. Following postdoctoral and faculty appointments in the biophysics division at Johns Hopkins, he moved to Los Alamos in 1983. From 1984 to 1989 Moyzis led the Laboratory Genetics Group, taking his current position as center director in 1989.

Maynard V. Olson earned his B.S. in chemistry from the California Institute of Technology in 1965 and his Ph.D. in chemistry from Stanford University in 1970. His doctoral thesis concentrated on physical inorganic chemistry. Olson joined the faculty of Dartmouth College as an Assistant Professor of Chemistry in 1969. He spent one year on sabbatical leave at the University of Washington and returned in 1976 as a Research Associate in the genetics laboratory of Benjamin Hall. In 1979 he joined the Department of Genetics at Washington University, where he is now a Professor of Genetics. In 1989 Olson became an investigator at the Howard Hughes Medical Institute at Washington University. His research specialties are in yeast and human genetics with an emphasis on the long-range organization of eukaryotic genomes and the structure and function of eukaryotic chromosomes. Olson has pioneered techniques for constructing physical maps, and in 1991 he completed a high-resolution physical map of the genome of the yeast *Saccharomyces cerevisiae*. He and his coworkers at Washington University developed the technology for cloning large DNA inserts in yeast artificial chromosomes (YACs). Olson is a member of the National Institutes of Health Program Advisory Committee on the Human Genome.

Nancy S. Wexler is an internationally respected authority on research into the genetic causes of human disease. In 1981 she began to study the world's largest known family with Huntington's disease, a family living along the shores of Lake Maracaibo in Venezuela. Over the years she and her colleagues have constructed a pedigree of over 12,000 people in the family and have collected blood samples from more than 3000 people. The samples led to the mapping of the Huntington's disease gene to the tip of human chromosome 4, which in turn led to the development of an effective presymptomatic test involving DNA markers that are tightly linked to the Huntington's disease gene. Wexler earned her A.B. from Radcliffe College in 1967 and her Ph.D. in Clinical Psychology from the University of Michigan in 1974. Wexler is President of the Hereditary Disease Foundation and Professor of Clinical Neuropsychology in the Departments of Neurology and Psychiatry of the College of Physicians and Surgeons at Columbia University. In addition to organizing and participating in many research collaborations, Wexler is a member of the Program Advisory Committee of NCHGR (National Center for Human Genome Research) and is chairperson of the Joint NIH/DOE ELSI (Ethical, Legal, and Social Implications of the Human Genome Project) Working Group.

Norton D. Zinder earned his A.B. from Columbia University in 1947 and his Ph.D. from University of Wisconsin in 1952. His first major discovery resulted from the attempt to induce matings between two strains of the bacterium *Salmonella typhimurium*. Zinder and Professor Joshua Lederberg found cell colonies that were the product of a hitherto unknown process (now called transduction) in which bacteriophages act as carriers of genetic material from one bacterial strain to another. After receiving his doctorate Zinder joined the faculty of Rockefeller University as an assistant professor. In 1960 Zinder and a graduate student, Timothy Loeb, discovered seven new viruses that infected only "male" strains of *E. coli*. The viruses, f1 through f7, proved to be unusual. The genetic material of f1 was found to contain a single strand of DNA and f2 through f7 were the first known RNA bacteriophages (phages whose genetic material is RNA). Zinder and his group demonstrated in 1962 that replication of an RNA phage is not dependent on DNA and that its RNA acts both as genetic material and as a template for directing protein synthesis. Zinder was appointed a professor at Rockefeller University in 1964 and the John D. Rockefeller Jr. Professor in 1977. His recent research emphasizes genetic recombination of the bacteriophage f1 and the physical mapping of its genome by means of restriction enzymes. In addition, he has conducted extensive nucleotide sequence analyses of messenger RNA from both prokaryotes and eukaryotes. An active editor, author, and spokesman on the responsibilities of scientists, Zinder has chaired many scientific committees and advisory panels, including a committee that evaluated the Virus Cancer Program of the National Cancer Institute, the Section of Genetics, National Academy of Sciences (1979–1982), the NAS/NRC (BAST) Committee on the Disposal of Chemical Weapons (1982–1984), and the Committee of Industry-University Relationships (COGENE) of the International Council of Scientific Unions (1982–1984). He was one of the original members of the Committee on Recombinant DNA Molecules of the National Research Council of the National Academy of Sciences in 1974–1975. He received the Eli Lilly Award in Microbiology in 1962 and the National Academy of Sciences' United States Steel Foundation Award in molecular biology in 1966 "for the discovery of RNA phages and for the analysis of the mechanisms of their replication." In 1982 Zinder received the AAAS Award in Scientific Freedom and Responsibility. ■